🛒 E-Commerce Product Review Analysis (GenAI Project)

📌 Overview

E-Commerce Product Review Analysis is a Generative AI–driven Natural Language Processing (NLP) system designed to automatically analyze customer reviews from online marketplaces.
The project classifies reviews into **Positive, Neutral, or Negative** sentiments and extracts meaningful insights related to product quality, delivery, pricing, performance, and customer satisfaction.

This solution helps **brands, sellers, and e-commerce platforms** make faster, data-driven decisions by eliminating manual review analysis.

🎯 Problem Statement

E-commerce platforms receive **thousands of customer reviews daily**.
Manual analysis is:

- Time-consuming
- Inconsistent
- Unable to extract insights at scale

There is a need for an **automated, real-time sentiment analysis system** that converts raw text reviews into actionable intelligence.

✅ Objectives

- Automatically classify product reviews as **Positive, Negative, or Neutral**
- Enable **real-time review sentiment prediction**
- Identify **key product aspects** (battery, design, performance, shipping, etc.)
- Reduce manual workload for sellers and platforms
- Deploy a **simple web interface** for live predictions

💡 Proposed Solution

The system uses a **machine learning pipeline** consisting of:

- **Text Preprocessing** (cleaning, lemmatization, stopword removal)
- **TF-IDF Vectorization** (unigrams + bigrams)
- **Linear Support Vector Machine (SVM)** for sentiment classification
- **Flask Web Application** for real-time predictions

🧠 Technology Stack

🔹 Languages & Frameworks

- Python
- Flask

🔹 Libraries

- Scikit-Learn
- NLTK
- Pandas, NumPy
- Matplotlib, Seaborn
- WordCloud

🔹 Tools

- Google Colab

📊 Dataset

- **Source:** Amazon Product Reviews
- **Total Reviews:** ~20,063
- **Columns:**
    - Review Title
    - Review Text
    - Rating
    - Sentiment Label

A smaller **sample dataset** was used for demo and deployment testing.

⚙️ Methodology

🔹 Data Preprocessing

- Lowercasing
- Punctuation & special character removal
- Stopword removal
- Lemmatization
- Duplicate & missing value handling

🔹 Exploratory Data Analysis (EDA)

- Word clouds (positive & negative)
- Review length distribution
- Sentiment distribution
- Rating vs sentiment analysis
- Bigram & keyword frequency analysis

🔹 Model Building

- TF-IDF feature extraction
- Trained multiple models:
    - Logistic Regression
    - Random Forest
    - **Support Vector Machine (Best Model)**

📈 Results & Performance

- **Best Model:** Linear SVM
- **Accuracy:** ~87.5%
- **Features:** TF-IDF Bigram Vectors

🔑 Key Insights

- Positive reviews dominate electronics products
- Frequent issues detected:
    - Shipping delays

    - Battery complaints
- Appreciated aspects:
    - Design
    - Performance

🌐 Web Application Workflow

1. User enters a product review
2. Backend preprocesses the text
3. TF-IDF vectorization applied
4. SVM model predicts sentiment
5. Output displayed with:
    - Sentiment label
    - Confidence score
    - Cleaned review text

🚀 How to Run the Project

🔹 Prerequisites

pip install numpy pandas nltk scikit-learn flask matplotlib seaborn wordcloud tqdm

🔹 Run the Application

python app.py

Open browser and go to:

http://127.0.0.1:5000/

🔮 Future Enhancements

- Transformer models (BERT, RoBERTa, DistilBERT)
- Aspect-based sentiment analysis
- REST API / microservice deployment
- Seller analytics dashboard

- Multilingual review support
- Fake review detection

🏁 Conclusion

This project demonstrates how **GenAI, NLP, and Machine Learning** can be effectively combined to analyze customer feedback at scale.
By automating sentiment classification and extracting meaningful insights, the system empowers sellers and platforms to improve **product quality, customer satisfaction, and decision-making efficiency**.

👥 Team Members

- Abhigna Rao Lingala
- M.Likith Rajesh
- E.Anil
- R.Nagesh
- B.Raja Vardhan

Thank you